CAN KNOWLEDGE IMPROVE POPULATION FORECASTS AT SUBCOUNTY LEVELS?*

GUANGQING CHI

Recent developments in urban and regional planning require more accurate population forecasts at subcounty levels, as well as a consideration of interactions among population growth, traffic flow, land use, and environmental impacts. However, the extrapolation methods, currently the most often used demographic forecasting techniques for subcounty areas, cannot meet the demand. This study tests a knowledge-based regression approach, which has been successfully used for forecasts at the national level, for subcounty population forecasting. In particular, this study applies four regression models that incorporate demographic characteristics, socioeconomic conditions, transportation accessibility, natural amenities, and land development to examine the population change since 1970 and to prepare the 1990-based forecast of year 2000 population at the minor civil division level in Wisconsin. The findings indicate that this approach does not outperform the extrapolation projections. Although the regression methods produce more precise projections, the least biased projections are often generated by one of the extrapolation techniques. The performance of the knowledge-based regression methods is discounted at subcounty levels by temporal instability and the scale effect. The regression coefficients exhibit a statistically significant level of temporal instability across the estimation and projection periods and tend to change more rapidly at finer geographic scales.

Population forecasting[1] at subcounty levels provides important information to local governments, businesses, and academics for various purposes. However, most of the subcounty forecasting methodologies currently viewed as state-of-the-art are projection techniques that were developed prior to 1960, although billions of dollars are allocated annually in public programs based on population forecasts (Smith, Nogel, and Cody 2002). Some recently enacted legislation demands innovations in subcounty population forecasting. For example, the "Smart Growth" law enacted in many states calls for more accurate forecasts. The Transportation Equity Act for the 21st Century of 1998 and the Clean Air Act Amendments of 1990 require transportation planners and environmental analysts to consider the interactions among traffic flows, land use, population growth, and environmental impacts (Smith, Tayman, and Swanson 2001). The existing subcounty extrapolation techniques are incapable of both forecasting population and estimating the relationships among these elements.

The multiple regression approach, which takes into account the influential and covariated factors of population change, could potentially address the policy demand. Nevertheless, no evidence supports the idea that current regression approaches produce more accurate population forecasts than the extrapolation projection techniques for subcounty areas. One possible reason regression approaches might not yield better forecasts is that they do not consider enough knowledge—many nondemographic factors from other disciplines that are not typically involved in formal population forecasting efforts

*Guangqing Chi, Department of Sociology and Social Science Research Center, Mississippi State University, P.O. Box C, Mississippi State, MS 39762; e-mail: gchi@soc.msstate.edu. Earlier versions of this article were presented at the annual meeting of the Southern Demographic Association, October 12–13, 2007, Birmingham, AL; and at the annual meeting of the Population Association of America, March 30–April 1, 2006, Los Angeles, CA. The author is grateful to Paul R. Voss for his guidance with this research and his insightful suggestions on earlier versions of the article. Appreciation is extended to Stephen J. Ventura, Isaac W. Eberstein, Patricia Dill, Glenn D. Deane, Deborah Balk, *Demography* editors, and four anonymous reviewers for their many helpful comments. The research was supported in part by faculty startup grants from the Office of Research and Economic Development of Mississippi State University and the Wisconsin Agricultural Experiment Station (Hatch project no. WIS04536).

1. A projection embodies one or more assumptions, and a forecast is a projection that is most likely to occur based on judgments. Nevertheless, "forecast" and "projection" are used interchangeably in this article.

can have significant effects on population change. In the context of population forecasts, knowledge refers to the relationships between the population and its demographic characteristics, socioeconomic phenomena, environment, energy, agriculture, legislation, and other nondemographic factors.

At the national level, numerous studies have succeeded in using the knowledge-based regression approach for population forecasting[2] (e.g., Ahlburg 1987a, 1987b; Forrester 1971; Lutz 1994; Meadows, Meadows, and Randers 1992; Meadows et al. 1972; Sanderson 1995). For instance, the World3 model, one of the most recognized and complex models, partially attempted to project the future world population after examining the dynamic relationships between population, agricultural production, natural resources, industrial production, and pollution. The World3 model produced a better 25-year world population forecast than the United Nations did (Sanderson 1998). In another example, Wheeler (1984) utilized the relationships between economics, human resources, and demographics to project population levels for developing countries. His forecasts also outperformed the United Nations' forecasts, which were based on simple demographic characteristics. Thus, Sanderson (1998) claimed that more knowledge can improve population forecasts. More recently, Lutz and Goldstein (2004) advocated the incorporation of substantive knowledge into formal population forecasting models.

These forecasts were conducted at the national level. Can knowledge also improve population forecasts for subcounty areas? In this study, more knowledge, especially that of nondemographic factors, is utilized to see whether these factors can improve subcounty population forecasting. The factors of socioeconomic conditions, transportation accessibility, natural amenities, and geophysical limitations will be integrated into subcounty population forecasting.

This article is organized into five additional sections. This introduction is followed by a review of regression approaches for small-area population forecasting. The Data section describes my research data, unit of analysis, and research variables. The Methods and subsequent Results sections employ four regression models to project 2000 populations at the minor civil division (MCD) level in Wisconsin, evaluate the four regression models and projections, and compare them with four extrapolation projections. Finally, I close this article with a concluding Summary and Discussion section.

PRIOR RESEARCH

The Regression Approach for Subcounty Population Forecasting

The familiar multiple regression models have been used in the production of population forecasts for more than 50 years (Schmitt 1953, 1954). Although Stanbery (1952) did not mention regression-based forecasts in his early "guide book" for population forecasting for small areas and communities, Pittenger (1976:68–77) devoted considerable attention to the matter in his comprehensive review of population projection models nearly a quarter-century later. Smith, Tayman, and Swanson (2001) dedicated two chapters to structural modeling in their recent and well-received comprehensive overview on the topic. Of particular relevance to this study is Chapter 9, which discusses economic-demographic structural models. More recently, Alho and Spencer (2005) discussed the regression approach in a chapter of their book as well. Yet most applied demographers producing population forecasts for small areas have largely ignored regression approaches. One justifiable reason for this appears to be that, thus far, no multiple regression models produce more accurate subcounty population forecasts than can be achieved by much simpler extrapolation

2. Unsurprisingly, some studies (e.g., Murdock et al. 1984; Smith and Sincich 1992) suggest that the regression approach does not outperform existing extrapolation methods at the national level.

techniques. The reason may be because for small subcounty areas (especially in sparsely settled rural landscapes), nondemographic factors, which assume a level of importance greater than whatever demographic forces appear to be at work, generally are ignored in traditional forecasting methodologies. “[P]opulation forecasts, like forecasts in sociology in general, tend to be self-contained; they use available population data and not much else” (Keyfitz 1982:730). Although existing regression approaches for population forecasting generally do consider numerous factors in explaining population change, these factors tend to be chosen by an unnecessarily narrow demographic perspective rather than a perspective informed by other theories and potential data sets.

Therefore, a further look at the multiple regression approach for small-area population forecasts is required, and this approach can be improved by examining, holistically rather than partially, the relationships between population change and relevant nondemographic factors from other disciplines not typically involved in formal population forecasting efforts. These population-related disciplines include human ecology, population geography, regional economics, environmental sociology, and urban and regional planning. Each of these disciplines has its strengths in interpreting and modeling population change. For example, human ecologists are interested in structural and organizational aspects of population (Frisbie and Poston 1975; Poston and Frisbie 1998); population geographers emphasize spatial regularities and processes in explaining and forecasting population patterns (Trewartha 1953); environmental sociologists (e.g., Commoner 1972; Schnaiberg and Gould 1994) and neo-Marxists (e.g., O’Connor 1989) emphasize political economy, class, and inequity in determining population change; some regional scientists (e.g., Deller et al. 2001) and rural demographers (e.g., Fuguitt and Brown 1990; Johnson 1999) emphasize the role of natural amenities in recent population redistribution processes; and transportation planners often utilize the relationship between transportation and population change for shaping land use patterns (Chi, Voss, and Deller 2006). These disciplines provide fresh views on population change, from which we can derive influential factors of population change and use them for small-area population forecasting. An exhaustive literature review of their theoretical works and empirical studies results in approximately 70 variables that are relevant to population change. The selected factors for this study are addressed in the Data section.

Exclusions of Alternative Approaches

I deliberately exclude five types of models from the regression forecasting specification I present: time-series models, postcensal population estimation models, conditional probabilistic models, integrated land use models, and population estimation and forecasting by grid cells. The time-series methods are based on trend modeling (linear or quadratic regressions fit on a historical time series), adaptive smoothing, and Box-Jenkins autoregressive integrated moving average (ARIMA) modeling (Box and Jenkins 1976). These methods have been used for population forecasting at national, state, and county levels with very mixed results (Alho and Spencer 1997; Land 1986; Land and Cantor 1983; Pflaumer 1992; Saboia 1974; Tayman, Smith, and Lin 2007). For subcounty areas, however, the time-series models are seldom adopted due to the lack of an appropriate time series.

The postcensal population estimation models (e.g., the ratio-correlation method) rely on contemporaneous systematic indicators for the regression-based estimate. Such indicators include tax returns, voter registration, school enrollment, telephone installations, utility meter connections, occupancy permits issued, and motor vehicle licenses. The symptomatic data are usually maintained by various public agencies for their own purposes and are more subject to errors (Greenberg, Krueckeberg, and Michaelson 1978). The literature covering this regression approach to postcensal estimation is large (e.g., Espenshade and Tayman 1982; Swanson and Beck 1994).

The conditional probabilistic forecasting approach is a competitive alternative to the regression approach. The 1990s and 2000s have seen an increasing development in probabilistic population forecasting, as evidenced by numerous journal articles and three special journal issues of the *International Journal of Forecasting* (guest edited by Ahlburg and Land 1992), the *Population and Development Review* (guest edited by Lutz, Vaupel, and Ahlburg 1998), and the *International Statistical Review* (guest edited by Lutz and Goldstein 2004). The recent conditional probabilistic forecasting approaches combine the advantages of the probabilistic approach, which quantifies the uncertainty range, with the benefits of scenario analysis, which is essential for policy making and measures the sensitivity of specific alternatives' sequences (e.g., Alho 1997; O'Neill 2004; Sanderson et al. 2004). Nevertheless, this approach is noted for being mechanistic and ignoring the environmental, geophysical, and transportation knowledge about the determinants of population change (Lutz and Goldstein 2004).

Integrated land use models can be very useful for small-area population forecasting. These models integrate demographic, socioeconomic, transportation, land use, and environmental components and use spatial data analysis and statistics techniques to simulate their spatiotemporal dynamics. These models adopt synthetic and interdisciplinary approaches as this study does, but the former is much more advanced and powerful than the latter. The widely used models include the classic Lowry (1964) model, the Garin-Lowry model (Garin 1966), the disaggregated residential allocation model (DRAM), the employment allocation model (EMPAL; Putman 1991), optimization models (e.g., Southworth 1995), land pricing models (e.g., Anas 1992), spatial economic models (e.g., Hunt and Simmons 1993), land-use change models (see Agarwal et al. 2002 for a thorough review of these models), and the recently developed land-use evolution and impact assessment model (LEAM; Deal et al. 2005). These models receive increasingly substantial demands because they can produce useful information about the complex interactions among the components and enable planners and decision makers to examine "what-if" planning scenarios (Tayman 1996). However, these models require high data capacity and well-grounded expertise in statistics and geographic information systems (GIS).

Population estimation and forecasting by grid cells has been traditionally used by geographers and recently utilized by demographers. Disaggregation (or interpolation) techniques play a vital role in this approach. The disaggregation techniques include inverse distance weighting (Bracken and Martin 1989), kriging (Cressie 1993), areal weighting (Flowerdew and Green 1992), smoothing (Tobler 1979), and the ancillary weighting based on local streets (e.g., Reibel and Bufalino 2005) or remote sensing images (e.g., Cowen and Jensen 1998; Langford, Maguire, and Unwin 1991). The recent efforts for forecasting global environmental change produce population forecasting by grid cells as evidenced in a special issue of *Technological Forecasting and Social Change* (guest edited by Riahi and Nakicenovic 2007). The recent work for global population estimation by grid cells includes the Gridded Population of the World (GPW), developed by the Center for International Earth Science Information Network (CIESIN) at Columbia University, and the LandScan, developed by the Geographic Information Science and Technology Group of the Oak Ridge National Laboratory. The major advantages of using this approach for subcounty population forecasting are that the remote sensing data are more timely than census data, and environmental and geophysical effects on population change at smaller areas are arguably better captured and modeled. However, similar to the integrated land-use models, this approach is expensive to develop and implement.

DATA

In this study, I examine whether the consideration of nondemographic factors can improve population forecasting for subcounty areas. Particularly, I focus on the state of Wisconsin as the research case and examine population change from 1970 to 2000 in Wisconsin at the MCD level.

The data used in this study come from a variety of sources, both primary and secondary. Population data are from decennial censuses from 1970 to 2000, and commercial reworking of these data is made available on the Geolytics Census CD. Demographic and socioeconomic data are acquired from the U.S. Census Bureau, the Federal Bureau of Investigation, and the State of Wisconsin Blue Books. Transportation infrastructure data are provided by the National Atlas of the United States, the Wisconsin Department of Transportation, the Wisconsin Bureau of Aeronautics, and the Department of Civil and Environmental Engineering of the University of Wisconsin–Madison. The data of geophysical factors and natural amenity characteristics come from the U.S. Geological Survey, the Wisconsin Department of Natural Resources, and the Environmental Remote Sensing Center and the Land Information and Computer Graphics Facility of the University of Wisconsin–Madison.

Unit of Analysis

This study is conducted at the MCD level. Wisconsin is a "strong MCD" state, and its MCDs—towns, cities, and villages—are functioning governmental units (with elected officials who provide services and raise revenues). The MCD geography consists of non-nested, mutually exclusive, and exhaustive political territories. In most parts of the state, census tracts have an average size similar to MCDs and provide an alternative unit of analysis. However, census tracts are geographic units delineated by the Census Bureau only for the purpose of counting the population, and tracts have no political or social meaning. In contrast, population projections at the MCD level have more "consumers," considering that governmental agencies are the largest users of demographic forecasts. In other words, the great advantage of using MCDs is their relevance to planning and public policy-making.

MCD boundaries are not stable over time. Boundaries change, new MCDs emerge, old MCDs disappear, names change, and status in the geographic hierarchy shifts (e.g., towns become villages, villages become cities). In order to adjust the data for these changes, I applied three rules: new MCDs must be merged into the original MCDs from which they emerged; disappearing MCD problems can be solved by dissolving the original MCDs into their current "home" MCDs; and occasionally, several distinct MCDs must be dissolved into one super-MCD in order to establish a consistent data set over time. The final analytical data set contains 1,837 MCDs with an average size of 29.56 square miles.

Explanatory Variables of Population Change

From the 70 variables derived from population-change-related disciplines, I selected 38 variables based on data availability to examine their relationships with population change and to project 2000 population. The variables are within the broad realms of demographic characteristics, socioeconomic conditions, physical infrastructure, environmental and geophysical factors, cultural resources, and potential legal constraints. Demographic and socioeconomic factors, which are familiar to demographers, are outlined only briefly here.

Demographic characteristics. It has long been understood that demographic characteristics of a population are important determinants of population change, and demographic characteristics should always be considered in any population forecast. The most important demographic characteristics that affect subcounty population growth are population density, age structure, racial and ethnic composition, institutional populations, educational attainment, nonmovers, female-headed families with children, and sustenance organization (Browning and Gibbs 1971; Friedman and Lichter 1998; Frisbie and Poston 1975; Humphrey 1980; Johnson and Purdy 1980; Lutz 1994; Mincer 1978).

Socioeconomic conditions. Socioeconomic conditions known to have important impacts on population change include employment opportunities, crime rate, school performance, income growth and distribution, public infrastructure, new housing, county seat status, real estate value, and local efforts to expand services (Carlino and Mills 1987; Clark and Murphy 1996; Deller et al. 2001; Hulten and Schwab 1984; Morrison and Schwartz 1996).

Transportation accessibility. Transportation accessibility is important for local economic growth and development, as well as associated population growth. Transportation accessibility can influence population change indirectly through economic growth or decline, employment change, altered social structure, and environmental change. There are many theoretical and empirical works on this topic in the fields of regional economics, transportation planning, rural sociology, and demography (for a summary of the literature, see Bhatta and Drennan 2003; Boarnet 1997; and Chi et al. 2006). Regional economic theories are especially strong in explaining the effects of transportation infrastructure on economic and population growth: the neoclassical growth theory is insightful in explaining and predicting metropolitan development after the transportation network is built, the growth pole theory is useful for forecasting population change from the standpoint of decision makers, and the location theory is strong in interpreting geographic distribution of population. Moreover, some rural demographers have studied population redistribution through residential preference and have found that migrants prefer locations somewhat rural or truly "sub"-urban within commutable distance of large cities (e.g., Brown et al. 1997; Fuguitt and Brown 1990; Zuiches and Rieger 1978). In this study, five variables represent transportation accessibility: proximity to central cities, accessibility to highways, accessibility to airports, highway infrastructure, and travel time to work.

Natural amenities. Environmental and natural resource characteristics are known to influence population growth. Disamenities (negative influences on population) include landfills and other noxious sites, resource extractions, and propensity to natural disasters. In recent decades, natural resource characteristics such as water features, terrain relief (e.g., viewsheds), and landscape aesthetics (e.g., regional land use and cover) have been viewed as influences on population change mainly through the role of natural amenities, which are seen as the principal contributor of the post-1970 "turnaround migration" in rural America (Brown et al. 1997; Frey 1987; Fuguitt 1985; Fuguitt and Brown 1990; Johnson 1999; Johnson and Purdy 1980). Equilibrium theory argues that the main determinants of migration come from differences in amenities rather than differences in economic opportunities (Graves and Linneman 1979). The life cycle literature suggests that amenity factors become more important as people become older (Clark and Hunter 1992; Humphrey 1980). Some regional economists see natural amenities as latent regional factor inputs to the local production of goods and services (English, Marcouiller, and Cordell 2000; Graves 1983; Knapp and Graves 1989; Porell 1982). They argue that natural amenities play a significant role in affecting economic development and migration. So-called growth engines in rural areas increasingly are less dependent upon traditional tangible factor inputs (land, labor, and capital) and more dependent upon latent factor inputs (such as amenity-based goods and services; Marcouiller, Kim, and Deller 2004). In this study, I use five variables of natural amenities: the proportion of forestry areas, the proportion of water areas, lengths of lakeshore/riverbank/coastline adjusted by the MCD's area,[3] total area of golf courses adjusted by the MCD's area, and the proportion of areas with slope between 12.5% and 20%.

Land development. Population growth is limited by the potential for land conversion and development. The land developability of a region is determined by its geophysical characteristics (water, wetlands, slope, and tax-exempt lands), developed lands (existing residential, commercial, and industrial developments, as well as transportation infrastructure), cultural and aesthetical resources, and legal constraints (including land use planning

3. The adjustment is based on the logic of shape analysis, which is basic to spatial structure of landscape elements. Readers who are interested in this topic should refer to Dryden and Mardia (1998).

legislation and programs such as comprehensive plans, "smart growth" laws, zoning ordinances, and farmland protection programs; and environmental regulations such as the Clean Water Act, shoreland and wetland zoning, and others). The existing literature includes developable lands (Cowen and Jensen 1998), qualitative environmental corridors (Lewis 1996), quantitative environmental corridors (Cardille, Ventura, and Turner 2001), and the growth management factors (Land Information & Computer Graphics Facility 2000).

Different from the four categories of variables addressed above, land development is best represented by an index rather than several variables. Environmental analysts often employ the ModelBuilder™ function of ArcGIS (ESRI 2000) to study the interactions between environment, population, land use, and legal constraints at fine pixel sizes. Although demographers may be interested in borrowing this approach to study population, the fact that population data are aggregated at rather coarse sizes imposes difficulties in taking into account environmental variables that generally can be studied usefully only at very fine data resolution. In this study, I employ the ModelBuilder function to generate a developability index that refers to the potential for land conversion and development. The general idea is to identify undevelopable lands at the pixel level, and then aggregate these to the MCD level for which the developability index is produced. The ModelBuilder function is first used to overlay the data layers of the variables (water, wetland, slope, tax-exempt lands, and built-up lands) and create one layer representing undevelopable lands for Wisconsin. This layer is then intersected with a geographic MCD layer to create a layer that contains the information for undevelopable lands at the MCD level. Based on that, the proportion of undevelopable land for each MCD is calculated, and the developability index is generated by subtracting the proportion of undevelopable land from 1.

Existing regression models for subcounty population forecasting are almost entirely sociodemographic. Their major shortcoming is that they generally ignore nondemographic factors that could be drawn from the contextual region for which the population forecasts are made. Population growth or decline has causes and consequences tied closely to levels of transportation accessibility as well as the nature of the surrounding natural environment. I argue that they can and should be incorporated into regression models for subcounty population forecasting.

METHODS

I use these variables (demographic and socioeconomic factors, transportation accessibility, natural amenities, and land development) to forecast the 2000 population based on their estimated relationships with population change from 1980 to 1990. First, I specify four regression models to estimate the relationships between the explanatory variables in 1980 and population change from 1980 to 1990. Second, I use the estimated relationships to project 2000 population. Third, I assess the four regression models on the basis of model diagnostics and against the actual 2000 population. Finally, I compare the four regression projections with four extrapolation projections to test the performance of the proposed knowledge-based approach.

Regressions and Projections

For the regression approach, the general assumption is that the effects of relevant factors on population change are constant over time. Thus, historical data can be used to estimate these effects, and then these effects (via the estimated parameters) can be applied to project future population. The first step is to use a multivariate linear regression model to build relationships between population change and relevant covariates. The regression models are specified in four ways by varying the representation of the dependent variable (population growth rate) and by including the constant or not (Eq. (1)).

$$\text{Population growth rate}_{1980\text{–}1990} = (\text{Population growth rate}_{1970\text{–}1980} + X_{1980}) \times \beta + \varepsilon, \quad (1)$$

where X_{1980} represents the independent variables in 1980, and β represents the corresponding coefficients expressing marginal relationships with the dependent variable.

Model 1 expresses the population growth rate as population change over the earlier census population (i.e., the difference of 1990 and 1980 populations over 1980 population) and includes the constant term. The explanatory variables comprise 34 individual variables, as listed in the previous section and displayed in Table 1. Model 2 is identical to Model 1 except that Model 2 excludes the constant term. Model 3 expresses the population growth rate as the natural log of the later census population over the earlier census population (i.e., 1990 population over 1980 population) and includes the constant term. Model 4 is identical to Model 3 except that the former excludes the constant term. Model 4 is my preferred model because the log transformation can achieve the desired bell-shaped distribution and better linearity with the independent variables, and the exclusion of the constant term can eliminate the disturbance of the change of population redistribution processes. The constant term represents the overall growth rate, which is identical for all MCDs. However, the overall growth rate does change from decade to decade, especially for this study. In Wisconsin, population distribution trends changed back to "renewed metropolitan growth" in the 1980s since the initial "nonmetropolitan turnaround" in the 1970s, and reversed once more to "rural rebound" in the 1990s (Johnson 1999). The intercept should be excluded from the regression model for population projection, and thus the overall growth rate is forced into the coefficients of the independent variables. However, this reasoning is yet to be tested in data analysis.

One or more of the independent variables may not be statistically significant in explaining population change. Moreover, the large number of variables may cause the multicollinearity problem and is not easily handled in the forecasting process. In the second step, insignificant ($p > .05$) independent variables are discarded from the regression model, which is run again until all retained independent variables are significant ($p \leq .05$). Finally, the variables derived from the second step are used to build a projection model (Eq. (2)) for each of the four regression models, and the variables are represented using data for the period 10 years later.

$$\widehat{\text{Population growth rate}_{1990\text{–}2000}} = (\text{Population growth rate}_{1980\text{–}1990} + X_{1990}) \times \hat{\beta}, \tag{2}$$

where X_{1990} represents the independent variables in 1990, and $\hat{\beta}$ represents the estimated parameters from the second step. From the left-hand term in Eq. (2), the 2000 population can be projected.

Model Comparisons

The four regression projection models are then evaluated from three perspectives (as proposed by Mandell and Tayman 1982): regression diagnostics, projection accuracy, and coefficient drift. In terms of regression diagnostics, this study assesses goodness-of-fit, multicollinearity, normality, and heteroskedasticity. Akaike's Information Criterion (AIC) is used to measure the fit of the model to the data but penalize models that are overly complex.[4] Models having a smaller AIC are considered the better models in the sense of model fitting balanced with model parsimony (Chi and Zhu 2008). The multicollinearity problem is diagnosed by a multicollinearity condition number, which is not a test statistic but a diagnostic of the regression stability due to multicollinearity (Anselin 2005). An indicator

4. Likelihood ratio tests can be performed to compare models that are nested (i.e., one simpler model can be reduced from the other more complex model by constraining certain parameters in the complex model). If two models are not nested, the AIC is often used. The use of R^2 and adjusted R^2 for comparing models is problematic and statistically irrelevant because the models have a different number of parameters and different dependent variables (Burnham and Anderson 2002).

over 30 typically suggests the multicollinearity problem. The normality of the errors is tested by the Jarque-Bera test (Bera and Jarque 1980), which has an asymptotic chi-square distribution with two degrees of freedom. Furthermore, the heteroskedasticity is diagnosed by the Breusch-Pagan (1979) test and Koenker-Bassett (1982) test.

Projection accuracy of the four regression models is evaluated by five quantitative measures: mean algebraic percentage error (MALPE), mean absolute percentage error (MAPE), root mean squared percentage error (RMSPE), median algebraic percentage error (MedALPE), and median absolute percentage error (MedAPE). All five measures focus on percentage errors rather than numerical errors, and thus they take into account the effects of population sizes. MALPE and MAPE are the two most commonly used measures (Tayman 1996; Tayman and Swanson 1996). The MALPE is a measure in which the positive and negative values can offset each other, so it is used mainly as a measure of bias. A positive MALPE indicates an overprojected population, and a negative MALPE indicates an underprojected population. In contrast, the MAPE is a measure in which positive and negative values do not offset each other. It indicates the average percent difference between the forecasted population and the actual population, regardless of over- or underprojection. The MAPE is used widely as a measure of forecast precision in evaluating population projections. The RMSPE is another measure of precision. The MedALPE and the MedAPE are utilized to measure the "typical" errors rather than the mean errors, and they ignore the effects of outliers (Smith et al. 2001).

A major problem with the regression approach for projection is the temporal instability of coefficients, also called *coefficient drift* (Mandell and Tayman 1982; Tayman and Schafer 1985). In using the regression forecasting approach, one assumes that the coefficients keep constant over time. However, this assumption is often violated, and the violation is often thought to be the primary error source of the regression forecasting approach (Namboodiri 1972). Here, I measure the extent and significance of coefficient drift in two ways. First, I rerun each of the four models using the variables 10 years later (1990–2000), and compare the estimated coefficients with those from Eq. (1). Second, I apply the Chow (1960) test to assess the significance of coefficient drift between the 1980–1990 and 1990–2000 regression specifications.

Projection Evaluations

I eventually compare these four regression projections with four simple but widely used methods: a 10-year-based linear extrapolation (Model A), a 20-year-based linear extrapolation (Model B), a 10-year-based exponential extrapolation (Model C), and a 20-year-based exponential extrapolation (Model D). Population projections based on some form of extrapolation of the past into the future are an established and fundamental population forecast technique used for small geographic areas in many states for many years. The extrapolation projections are reliable and widely used for population forecasting and are at least as accurate as more sophisticated models in projecting short- to medium-term total population (Smith 1987).

I also compare the four regression projections with the four extrapolation projections based on population size and growth, both by the MALPE and MAPE. The statistical quality of projection is affected by population size and growth rate—forecast accuracy increases as population growth rate (in absolute terms) decreases, and as population size increases until reaching a threshold level that varies with numerous factors (Smith 1987).

RESULTS

Estimation and Diagnostics of the Regression Models

In the initial step, four regression models are estimated to examine the covariates between population change from 1980 to 1990 and the independent variables. The dependent variable is expressed either as the ratio of the 1980–1990 population change over the 1980 population (Models 1 and 2), or as the natural log of the 1990 census population over the 1980 population (Models 3 and 4). The independent variables include population change from 1970 to 1980 and other explanatory variables (see Table 1). Models 1 and 3 include the constant term, while Models 2 and 4 exclude it. Models 3 and 4 have much lower AIC values than Models 1 and 2, indicating that the logarithm transformation, rather than the nonlogarithm form of the dependent variable, can help improve model fitting balanced with model parsimony. Models 2 and 4 have much smaller multicollinearity condition numbers than Models 1 and 3, respectively, suggesting that the exclusion of the constant term drastically reduces the multicollinearity. The Jarque-Bera test shows that Models 1 and 2 are much worse fits than Models 3 and 4 in terms of normality. The residuals of all four models are significantly heteroskedastic. The Breusch-Pagan test indicates that Model 3 and Model 4 have less heteroskedasticity than Models 1 and 2, but the Koenker-Bassett test indicates the opposite.

In an attempt to improve the initial regression models, I eliminate the insignificant independent variables to reduce the extent of multicollinearity in the regression models. Table 2 shows the variables that were retained for the refined models. All retained independent variables are significant in explaining population growth. The multicollinearity is reduced dramatically in all four models and is eliminated in Model 2 (with a multicollinearity condition number of 11.12) and Model 4 (with a multicollinearity condition number of 29.71). All other statistics (the AIC value, and Jarque-Bera, Breusch-Pagan, and Koenker-Bassett tests) further indicate improvements in the refined models. The comparison among the four models by these statistics does not change the conclusions from comparing the initial regression models. Overall, the logarithm representation of population growth rate achieves better model fitting balanced with model parsimony, and improves the normality of the model. The exclusion of the constant term remarkably reduces the multicollinearity and eliminates it in the refined models. Thus, Model 4 is the preferred regression model.

Temporal Instability of the Regression Coefficients

The regression coefficients exhibit significant temporal instability between the estimation and projection periods. Table 3 reports the coefficients of the four regression models, which I rerun using the variables in the 1990–2000 period. Whereas the coefficients for the proportion of housing units using public water and the lengths of lakeshore/riverbank/coastline remain relatively constant across the 1980–1990 and 1990–2000 models, the coefficients for all the other variables change drastically.

Nevertheless, we do not know whether the difference in coefficients is statistically significant. I use the Chow test to assess the coefficient drift; the results are reported in Table 4. For each of the four models, the F statistic is much higher than the corresponding critical F value at the .001 level of significance, confirming that the set of coefficients between 1980–1990 and 1990–2000 are statistically different.

The temporal instability of the regression coefficients eventually affects the accuracy of the knowledge-based regression forecasting approaches in this study. One possible reason for the significant existence of coefficient drift is that population redistribution patterns were different in the two decades. Wisconsin experienced “renewed metropolitan growth” in the 1980s and reversed to “rural rebound” in the 1990s (Johnson 1999). The two processes are driven by different determining factors, and thus the effects and significance of some relevant variables on population change differ considerably. Some important (or unimportant) explanatory variables in the 1980s may lose (or gain) their importance in the 1990s, and the positive (or negative) effects of some variables in the 1980s may become negative (or positive) in the 1990s. These changes undermine the effectiveness of the regression approach for forecasting.

Table 1. Coefficients of Initial Standard Regressions (1980–1990)

	Nonlogged Population Growth		Logged Population Growth	
Variable	With Constant, Model 1	Without Constant, Model 2	With Constant, Model 3	Without Constant, Model 4
Demographic Characteristics				
Population growth rate, 1970–1980	0.064*** (0.020)	0.064*** (0.020)	0.065** (0.024)	0.066** (0.024)
Population density in 1980	–6.95E–5** (2.21E–5)	–6.73E–5** (2.23E–5)	–5.54E–5** (1.86E–5)	–5.31E–5** (1.88E–5)
Proportion of young population (aged 12–18) in 1980	–1.078*** (0.181)	–1.034*** (0.174)	–1.040*** (0.179)	–0.994*** (0.172)
Proportion of old population (aged 65+) in 1980	–0.414*** (0.104)	–0.390*** (0.104)	–0.390*** (0.097)	–0.365*** (0.097)
Proportion of black population in 1980	0.190 (0.300)	0.244 (0.308)	0.121 (0.272)	0.177 (0.278)
Proportion of Hispanic population in 1980	0.150 (0.367)	0.220 (0.368)	0.144 (0.348)	0.217 (0.349)
Proportion of college population in 1980	–0.034 (0.135)	–0.020 (0.134)	–0.064 (0.118)	–0.048 (0.117)
Proportion of population (aged 25+) who finished high school in 1980	0.011 (0.051)	0.029 (0.050)	0.004 (0.048)	0.022 (0.047)
Proportion of population (aged 25+) with Bachelor's degree in 1980	–0.064 (0.077)	–0.084 (0.077)	–0.061 (0.073)	–0.082 (0.073)
Proportion of nonmovers (aged 5+) in 1980	–0.050 (0.034)	–0.048 (0.034)	–0.050 (0.032)	–0.048 (0.032)
Proportion of families headed by female with children younger than 18 in 1980	–0.097 (0.102)	–0.073 (0.101)	–0.115 (0.101)	–0.090 (0.100)
Proportion of seasonal housing units in 1980	0.018 (0.037)	0.020 (0.037)	0.013 (0.035)	0.014 (0.035)
Proportion of workers in retail industry in 1980	0.067 (0.088)	0.056 (0.088)	0.046 (0.082)	0.034 (0.081)
Proportion of workers in agricultural industry in 1980	–0.027 (0.037)	–0.039 (0.036)	–0.037 (0.037)	–0.049 (0.036)
Socioeconomic Conditions				
Employment rate in 1980	0.148 (0.104)	0.185 (0.102)	0.166 (0.097)	0.205* (0.095)
Median household income in 1980	–1.39E–6 (1.73E–6)	–9.16E–7 (1.69E–6)	–2.19E–6 (1.61E–6)	–1.71E–6 (1.58E–6)
Crime rate in 1980	0.554** (0.210)	–0.041 (0.102)	–0.590** (0.195)	–0.058 (0.096)
Proportion of housing units using public water in 1980	0.055*** (0.017)	0.055** (0.017)	0.049*** (0.015)	0.048*** (0.015)
Proportion of new housing units (40 or fewer years) in 1980	0.108*** (0.029)	0.116*** (0.029)	0.100*** (0.029)	0.109*** (0.029)
Median house value in 1980	9.34E–7* (4.42E–7)	9.03E–7* (4.41E–7)	1.12E–6** (4.21E–7)	1.09E–6** (4.22E–7)

(continued)

(Table 1, continued)

Variable	Nonlogged Population Growth: With Constant, Model 1	Nonlogged Population Growth: Without Constant, Model 2	Logged Population Growth: With Constant, Model 3	Logged Population Growth: Without Constant, Model 4
Socioeconomic Conditions (cont.)				
Proportion of workers using public transportation to work in 1980	0.032 (0.221)	0.066 (0.218)	–0.024 (0.210)	0.012 (0.207)
Having urban buses or not in 1979	0.020 (0.021)	0.020 (0.021)	0.023 (0.020)	0.023 (0.019)
County seat status	0.036*** (0.011)	0.035*** (0.011)	0.037*** (0.010)	0.036*** (0.010)
Transportation Accessibility				
Inverse distance to the centroid of central cities	–38.227 (23.118)	–25.060 (20.915)	–38.805 (23.074)	–25.171 (20.169)
Inverse distance from the centroid of an MCD to its nearest major airport	53.352 (53.032)	82.310 (52.562)	49.097 (48.532)	79.093 (47.912)
Inverse distance to interchange of interstate highways	0.012 (0.042)	0.014 (0.040)	0.021 (0.043)	0.023 (0.040)
Highway lengths (adjusted by lengths over the square root of the area of an MCD)	–3.96E–4 (0.002)	–6.83E–4 (0.002)	–3.48E–4 (0.002)	–6.45E–4 (0.002)
Journey to work (proportion of workers traveling 30 or fewer minutes to work in 1980)	–0.028 (0.024)	–0.030 (0.025)	–0.028 (0.023)	–0.029 (0.023)
Natural Amenities				
Proportion of forestry areas	0.015 (0.021)	0.021 (0.021)	0.017 (0.020)	0.023 (0.020)
Proportion of water areas	0.006 (0.047)	0.006 (0.047)	0.022 (0.045)	0.022 (0.045)
Lengths of lakeshore/riverbank/coastline adjusted by the MCD's area	–0.001*** (0.0004)	–0.001*** (0.0004)	–0.001*** (0.0004)	–0.001** (0.0004)
Golf courses	3.94E–7 (2.10E–7)	3.92E–7 (2.18E–7)	4.36E–7* (1.94E–7)	4.35E–7* (2.02E–7)
Proportion of areas with slope between 12.5% and 20%	0.024 (0.045)	0.024 (0.045)	0.019 (0.046)	0.018 (0.046)
Land Development				
Developability	0.036 (0.022)	0.041 (0.022)	0.038 (0.022)	0.044* (0.022)
Constant	0.566* (0.224)	—	0.586** (0.208)	—
Diagnostics				
AIC	–2,741.58	–2,736.48	–2,943.73	–2,937.23
Multicollinearity condition number	484.34	148.01	484.79	148.15
Jarque-Bera test	6,869.14***	6,662.55***	1,288.21***	1,295.47***
Breusch-Pagan test	320.13***	304.55***	221.56***	184.98***
Koenker-Bassett test	57.31**	55.17*	72.65***	60.52**

Note: Robust standard errors are in parentheses.

*$p \le .05$; **$p \le .01$; ***$p \le .001$

Table 2. Coefficients of Refined Standard Regressions (1980–1990)

	Nonlogged Population Growth		Logged Population Growth	
Variable	With Constant, Model 1	Without Constant, Model 2	With Constant, Model 3	Without Constant, Model 4
Explanatory Variables				
Population growth rate, 1970–1980	0.079*** (0.019)	0.082*** (0.019)	0.089*** (0.022)	0.089*** (0.022)
Population density in 1980	–8.26E–5*** (1.69E–5)	–7.8E–5*** (1.7E–5)	–7.2E–5*** (1.43E–5)	–6.9E–5*** (1.45E–5)
Proportion of young population (aged 12–18) in 1980	–1.135*** (0.181)	–0.670*** (0.082)	–1.102*** (0.180)	–1.072*** (0.157)
Proportion of old population (aged 65+) in 1980	–0.410*** (0.081)	–0.225*** (0.062)	–0.363*** (0.077)	–0.360*** (0.072)
Employment rate in 1980	—	—	—	0.119*** (0.033)
Crime rate in 1980	–0.551** (0.192)	—	–0.552** (0.179)	—
Median house value in 1980	7.77E–7*** (2.29E–7)	1.20E–6*** (2.2E–7)	7.42E–7*** (2.23E–7)	7.1E–7** (2.29E–7)
Proportion of new housing units (40 or fewer years) in 1980	0.115*** (0.021)	0.143*** (0.021)	0.102*** (0.021)	0.124*** (0.021)
Proportion of housing units using public water in 1980	0.051*** (0.014)	0.061*** (0.014)	0.045*** (0.012)	0.044*** (0.012)
County seat status	0.037*** (0.010)	0.035*** (0.010)	0.037*** (0.009)	0.037*** (0.009)
Lengths of lakeshore/riverbank/coastline adjusted by area	–0.001*** (3.68E–4)	–0.001** (3.49E–4)	–0.001*** (3.64E–4)	–0.002*** (0.0004)
Golf courses	—	—	4.20E–7** (1.83E–7)	4.4E–7* (1.92E–7)
Constant	0.666*** (0.192)	—	0.658*** (0.179)	—
Diagnostics				
AIC	–2,766.47	–2,743.65	–2,965.52	–2,960.69
Multicollinearity condition number	274.96	11.12	281.13	29.71
Jarque-Bera test	6,447.24***	6,099.55***	1,237.48***	1,236.44***
Breusch-Pagan test	224.90***	197.94***	138.20***	134.33***
Koenker-Bassett test	41.30***	37.07***	45.92***	44.64***

Note: Robust standard errors are in parentheses.

*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$

Comparison of Projection Accuracy

I evaluate the accuracy of the four regression forecasting models compared with that of four extrapolation projections based on the quantitative measures, population size, and population growth rate. First, the quantitative measures of MALPE, MAPE, RMSPE, MedALPE, and MedAPE do not suggest a strong preference for regression models or extrapolation projections (see Table 5). None of the regression models and extrapolation projections achieves higher accuracy than the others uniformly in terms of both bias and precision. Within the four regression models, Models 1 and 4 achieve higher accuracy than Models 2

Table 3. Coefficients of Standard Regressions (1990–2000)

Variable	Nonlogged Population Growth: With Constant, Model 1	Nonlogged Population Growth: Without Constant, Model 2	Logged Population Growth: With Constant, Model 3	Logged Population Growth: Without Constant, Model 4
Explanatory Variables				
Population growth rate, 1980–1990	0.225*** (0.049)	0.226*** (0.047)	0.157*** (0.041)	0.158*** (0.041)
Population density in 1990	–1.29E–4*** (2.90E–5)	–1.28E–4*** (2.64E–5)	–1.03E–4*** (2.19E–5)	–1.03E–4*** (2.06E–5)
Proportion of young population (aged 12–18) in 1990	–0.471** (0.182)	–0.519*** (0.118)	–0.393* (0.162)	–0.415** (0.156)
Proportion of old population (aged 65+) in 1990	–0.160 (0.104)	–0.182* (0.079)	–0.128 (0.083)	–0.139 (0.078)
Employment rate in 1990	—	—	—	–0.017 (0.037)
Crime rate in 1990	–0.076 (0.396)	—	0.024 (0.304)	—
Median house value in 1990	4.65E–7 (2.60E–7)	4.44E–7 (2.44E–7)	4.12E–7 (2.26E–7)	4.07E–7 (2.31E–7)
Proportion of new housing units (40 or fewer years) in 1990	0.330*** (0.035)	0.326*** (0.031)	0.295*** (0.030)	0.290*** (0.028)
Proportion of housing units using public water in 1990	0.055** (0.021)	0.053* (0.021)	0.043** (0.016)	0.043** (0.016)
County seat status	–0.022 (0.012)	–0.022 (0.012)	–0.011 (0.010)	–0.011 (0.010)
Lengths of lakeshore/riverbank/coastline adjusted by area	–0.001* (5.12E–4)	–0.001* (4.64E–4)	–7.79E–4 (4.35E–4)	–8.16E–4 (4.32E–4)
Golf courses	—	—	2.21E–8 (1.91E–7)	2.77E–8 (1.91E–7)
Constant	0.059 (0.395)	—	0.047 (0.300)	—
Diagnostics				
AIC	–1,281.54	–1,285.33	–2,034.41	–2,036.12
Multicollinearity condition number	347.41	10.71	354.06	24.13
Jarque-Bera test	41,561.69***	41,664.29***	2,345.54***	2,368.47***
Breusch-Pagan test	403.49***	294.04***	221.33***	218.40***
Koenker-Bassett test	32.56***	23.71**	59.06***	58.07***

Note: Robust standard errors are in parentheses.

$^{*}p \le .05$; $^{**}p \le .01$; $^{***}p \le .001$

and 3. Model 3 outperforms Model 2 in four of the five measures, and Model 4 outperforms Model 1 in three of five measures, although the advantage is only marginal. Within the four extrapolation projections, the two 20-year-based extrapolation projections outperform the two 10-year-based ones. The 10-year-based exponential extrapolation projection (Model C) slightly outperforms the 10-year-based linear extrapolation projection (Model A). The 20-year-based exponential extrapolation projection (Model D) produces less biased and

Table 4. Chow Test of Regression Stability

Model	Degrees of Freedom	Critical *F* Value (at the .001 level of significance)	*F* Statistic
Model 1	(11, 3,652)	2.85	26.65
Model 2	(9, 3,656)	3.11	33.77
Model 3	(12, 3,650)	2.75	23.57
Model 4	(11, 3,652)	2.85	24.11

less precise projections than the 20-year-based linear extrapolation projection (Model B). Among the eight methods, Model D produces the least biased projections, while Models 1 and 4 produce the most precise projections.

Second, the regression models and extrapolation projections are compared on the basis of population size in 2000 by MALPE and MAPE (Table 6). The results do not indicate a strong preference for the regression models or extrapolation projections. For MCDs with 250 and fewer persons (118 MCDs), all four regression methods outperform the extrapolation methods, with the exception that the least biased projection is produced by Model B. For MCDs with 251–2,000 persons (1,310 MCDs) and MCDs with 2,001–20,000 persons (372 MCDs), the regression models provide slightly more precise projections than the extrapolation methods (except Model B), but Models B and D offer the least biased projections. For MCDs with more than 20,000 persons (37 MCDs), the regressions are much less biased and marginally more precise than the extrapolation projections. Overall, regression methods outperform the extrapolations in MCDs with fewer than 250 or more than 20,000 persons in 2000, which account for only 8% of all MCDs. Regression methods are slightly more precise but more biased than the extrapolation methods in MCDs with 251–20,000 persons. Projection accuracy is similar across each population segment within the four regression models but varies dramatically within the four extrapolation projections.

Third, the regression models and extrapolation projections are compared on the basis of population growth rate from 1990 to 2000 by MALPE and MAPE (Table 7). Again, the results do not suggest a strong preference for the regression methods or extrapolation techniques. For MCDs losing population (386 MCDs), the regression methods produce more precise projections than the extrapolation methods, but the least biased projections are always produced by one of the extrapolation methods in all three population decline groups. For MCDs without population change in the 1990s (8 MCDs) and MCDs gaining less than

Table 5. Evaluating Population Projections by Quantitative Measures (percentages)

Model	MALPE	MAPE	RMSPE	MedALPE	MedAPE
Regression Methods					
Model 1	–4.85	10.49	14.62	–4.78	7.91
Model 2	–5.34	10.59	14.57	–5.34	8.11
Model 3	–5.35	10.65	14.74	–5.26	8.15
Model 4	–4.90	10.49	14.60	–4.77	7.98
Extrapolation Methods					
Model A (10-year-based linear)	–8.49	13.41	18.02	–8.39	10.56
Model B (20-year-based linear)	–3.86	10.70	14.92	–3.85	8.18
Model C (10-year-based exponential)	–7.17	13.04	17.70	–7.47	10.55
Model D (20-year-based exponential)	–1.19	11.32	16.74	–2.12	8.46

Table 6. Evaluating Population Projections by Population Size in 2000 (percentages)

	Population 250 or Less (118 MCDs)		Population 251 to 2,000 (1,310 MCDs)		Population 2,001 to 20,000 (372 MCDs)		Population 20,001 or More (37 MCDs)	
Model	MALPE	MAPE	MALPE	MAPE	MALPE	MAPE	MALPE	MAPE
Regression Methods								
Model 1	1.95	15.55	–5.15	10.08	–6.29	10.94	–1.51	4.39
Model 2	0.74	14.98	–5.66	10.25	–6.45	10.99	–2.31	4.54
Model 3	1.30	15.44	–5.73	10.30	–6.57	10.98	–0.83	4.61
Model 4	1.50	15.44	–5.26	10.12	–6.06	10.80	–0.91	4.55
Extrapolation Methods								
Model A	–6.84	21.80	–8.48	13.19	–9.46	12.25	–4.53	6.12
Model B	–0.39	16.79	–3.70	10.26	–5.39	10.79	–5.09	5.99
Model C	–3.92	20.74	–7.37	12.87	–7.86	11.93	–3.71	5.69
Model D	2.05	17.06	–1.57	10.70	–0.73	12.17	–2.63	6.47

5% population (283 MCDs), the regression models are generally substantially more precise and less biased than the extrapolation methods. For MCDs gaining 5%–10% population (299 MCDs), the regression models are more precise than the extrapolation methods, but Model D produces the least biased projections. For MCDs gaining 10% population and more (861 MCDs), Model B produces the most precise projection, and Model D generates the least biased projection.

Overall, the four regression methods produce more precise projections than the four extrapolation methods, but the least biased projections are often produced by one of the extrapolation methods. Thus, the findings do not support that more knowledge—especially the nondemographic factors such as transportation, natural environment, and land development—can help improve subcounty population forecasting.

SUMMARY AND DISCUSSION

This study attempts to provide an interdisciplinary, theoretically grounded approach to subcounty population forecasting. It builds on existing theoretical foundations that hypothesize strong correlations between population change and its relevant influential factors. Traditionally, neither mathematical nor applied demographers have considered nondemographic variables, and when applying existing multivariate regression approaches, they often choose variables within their disciplinary framework. My premise was that such an approach is insufficient for modeling population change. Although recognizing a variety of causes and contexts of population change, demographers, especially applied demographers, have simply not implemented a more holistic approach to their work. It is not surprising that their advanced methods do not outperform existing projection techniques for subcounty areas. I asserted that subcounty population forecasting should consider nondemographic factors from other disciplines not typically involved in formal population forecasting efforts.

However, the findings of this study do not support the premise that this knowledge-based approach outperforms existing extrapolation projection methods. Although the regression methods produce more precise projections than the extrapolation techniques, the least biased projections are often generated by one of the extrapolation techniques. This observation is consistent in comparisons by population size and growth rate.

Table 7. Evaluating Population Projections by Population Growth Rate (nonlogged form) From 1990 to 2000 (percentages)

Model	−10% or less (97 MCDs)		−10% to −5.01% (105 MCDs)		−5% to −0.01% (184 MCDs)		0% (8 MCDs)		0.01% to 4.99% (283 MCDs)		5% to 9.99% (299 MCDs)		10% or more (861 MCDs)	
	MALPE	MAPE	MALPE	MAPE	MALPE	MAPE	MALPE	MAPE	MALPE	MAPE	MALPE	MAPE	MALPE	MAPE
Regression Methods														
Model 1	24.41	24.41	8.80	9.27	5.15	6.01	1.92	3.98	0.20	3.87	−3.47	4.78	−14.15	14.24
Model 2	23.50	23.50	8.46	8.71	4.52	5.53	1.22	3.24	−0.40	3.80	−4.06	5.14	−14.51	14.63
Model 3	23.74	23.74	8.15	8.61	4.60	5.57	0.73	3.14	−0.33	3.92	−3.99	5.23	−14.58	14.67
Model 4	24.31	24.31	8.67	9.08	4.98	5.80	1.37	3.53	0.06	3.86	−3.52	4.94	−14.12	14.27
Extrapolation Methods														
Model A	21.08	25.28	2.97	9.58	−1.01	7.56	−17.09	18.13	−4.60	7.92	−7.18	9.17	−16.47	17.02
Model B	23.39	25.34	8.07	10.44	3.46	7.27	−2.76	13.24	0.03	6.47	−3.58	6.74	−11.33	12.56
Model C	23.99	26.34	4.58	9.06	−0.07	7.20	−10.14	11.22	−3.71	7.66	−6.30	9.08	−15.04	16.45
Model D	27.20	27.99	9.62	11.07	4.92	8.05	1.96	12.96	1.50	7.23	−2.18	7.56	−7.58	12.82

Population Growth Rate applies to all column groups.

Why doesn't more knowledge produce more accurate forecasting? It is a traditional perception that the more we know about our society and environment, the better we can predict the future. A quarter-century ago, Keyfitz (1982) discussed the dilemma and provided six possible reasons, one of which is related to this study: the issue of temporal instability. In using the regression approach for population forecasting, one assumes that the influential factors' effects on population change are consistent across the estimation and projection periods. However, this assumption can hardly hold. The population redistribution process has experienced different patterns and has been affected by various factors, such as those listed in Table 1. When the primary influential factors change, population redistribution patterns change. Thus, the statistical results of the influential factors' effects are different in different time periods. The performance of the regression methods is discounted, especially in this study, because the population redistribution process has experienced opposite patterns: "renewed metropolitan growth" in the estimation period and "rural rebound" in the projection period. The explanation of temporal instability is supported by the findings of this study. A statistically significant level of temporal instability of the regression coefficients exists in this study and greatly affects the regression projection accuracy.

The performance of the regression approach for population forecasting is further lessened at finer scales. That more knowledge does improve forecasts at the national level in many studies but does not at the MCD level in this study may be due to a phenomenon of "scale effect" (Fotheringham and Wong 1991). The scale effect refers to the fact that when the same data are aggregated at different scales, results of statistical analysis are disparate. A change of the influential factors can have greater impacts on population change at subcounty levels than that at county, state, and national levels. When the local changes are aggregated from finer scales to larger ones, the changes may cancel each other out. For example, the relocation of a factory from one MCD to another decreases population in the former and increases population in the latter, but the internal migration has no impact on total population in their home county and state. Migration is substantially more volatile than fertility and mortality at subcounty levels, and migration often accounts for a larger fraction of population change than it does at the national level (Keyfitz 1972).

Knowledge does not improve population forecasts at subcounty levels due to temporal instability and the scale effect. "Thus at the end of this lengthy search we are driven back to statistical and mathematical methods that in one form or another, since they do not depend on outside knowledge or relations beyond the demographic series themselves, can only be called extrapolatory. Pending the discovery of a truly behavioral way of estimating the future, we cannot afford to be ashamed of extrapolating the observed regularities of the past" (Keyfitz 1982:747).

This daunting message makes one wonder what the research and practice of subcounty population forecasting should focus on, since improving forecasting accuracy seems infeasible. Applied demographers are the producers of various population forecasting products. There are increasing demands for subcounty population forecasting from governmental agencies, businesses, academic units, and nonprofit organizations. Most users tend to accept whatever demographers provide and do not pay much attention to the quality of the products (Ahlburg and Lutz 1998). Applied demographers are in the position of evaluating the projections, not only by the total population size and growth rate, but also by age-race-sex specifications and other demographic characteristics. It is essential to know the qualities, strengths, and weaknesses of each projection method, as well as inform the users of this information (Smith and Tayman 2003).

Despite the findings that the regression approach does not outperform the extrapolation projection for subcounty areas, the former does bring a significant advantage to local planning and decision-making that cannot be provided by the latter. The effort toward using more knowledge for population forecasts has never ceased, and should not in the future. Decision-makers are often interested in the "what-if" scenarios (Land and Schneider 1987).

Some producers using fundamental population forecasting methods have to provide the analysis separately. The proposed knowledge-based regression method allows analysts to examine the variety of results based on some variables of the full set. The estimated relationships between population change and relevant factors can inform planners and decision-makers of the possible consequences of adopting a strategy, as well as suggest strategies to solve potential development problems. In addition, the projected populations and estimated relationships are useful to scientific research in the natural sciences because some of them require knowledge about population-geophysics dynamics in small areas.

REFERENCES

Agarwal, C., G.M. Green, J.M. Grove, T.P. Evans, and C.M. Schweik. 2002. "A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice." General Technical Report NE-297. USDA Forest Service Northeastern Research Station and Indiana University Center for the Study of Institutions, Populations, and Environmental Change.

Ahlburg, D.A. 1987a. "The Impact of Population Growth on Economic Growth in Developing Nations: The Evidence From Macroeconomic-Demographic Models." Pp. 492–522 in *Population Growth and Economic Development: Issues and Evidence,* edited by D.G. Johnson and R.D. Lee. Madison, WI: University of Wisconsin Press.

———. 1987b. "Modeling Macroeconomic-Demographic Linkages: A Study of Models of National and Regional Economies." Pp. 287–336 in *Forecasting in the Natural and Social Sciences,* edited by K. Land and S. Schneider. Boston: D. Reidel.

Ahlburg, D.A. and K.C. Land, Guest Eds. 1992. *Population Forecasting*, special issue of *International Journal of Forecasting* 8(3).

Ahlburg, D.A. and W. Lutz. 1998. "Introduction: The Need to Rethink Approaches to Population Forecasts." *Population and Development Review* 24(Suppl.):1–14.

Alho, J.M. 1997. "Scenarios, Uncertainty and Conditional Forecasts of the World Population." *Journal of the Royal Statistical Society Series A—Statistics in Society* 160:71–85.

Alho, J.M. and B.D. Spencer. 1997. "The Practical Specification of the Expected Error of Population Forecasts." *Journal of Official Statistics* 13:203–25.

———. 2005. *Statistical Demography and Forecasting*. New York: Springer.

Anas, A. 1992. *NYSIM (The New York Simulation Model): A Model for Cost-Benefit Analysis of Transportation Projects*. New York: Regional Plan Association.

Anselin, L. 2005. "Exploring Spatial Data With GeoDa™: A Workbook." Spatial Analysis Laboratory, University of Illinois at Urbana-Champaign.

Bera, A.K. and C.M. Jarque. 1980. "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals." *Economics Letters* 6:255–59.

Bhatta, S.D. and M.P. Drennan. 2003. "The Economic Benefits of Public Investment in Transportation—A Review of Recent Literature." *Journal of Planning Education and Research* 22:288–96.

Boarnet, M.G. 1997. "Highways and Economic Productivity: Interpreting Recent Evidence." *Journal of Planning Literature* 11:476–86.

Box, G.E.P. and G.M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Bracken, I. and D. Martin. 1989. "The Generation of Spatial Population Distributions From Census Centroid Data." *Environment and Planning A* 21:537–43.

Breusch, T.S. and A.R. Pagan. 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* 47:1287–94.

Brown, D.L., G.V. Fuguitt, T.B. Heaton, and S. Waseem. 1997. "Continuities in Size of Place Preferences in the United States, 1972–1992." *Rural Sociology* 62:408–28.

Browning, H.L. and J.P. Gibbs. 1971. "Intraindustry Division of Labor: The States of Mexico." *Demography* 8:233–45.

Burnham, K.P. and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.

Cardille, J.A., S.J. Ventura, and M.G. Turner. 2001. "Environmental and Social Factors Influencing Wildfires in the Upper Midwest, USA." *Ecological Applications* 11:111–27.
Carlino, G.A. and E.S. Mills. 1987. "The Determinants of County Growth." *Journal of Regional Science* 27:39–54.
Chi, G., P.R. Voss, and S.C. Deller. 2006. "Rethinking Highway Effects on Population Change." *Public Works Management and Policy* 11:18–32.
Chi, G. and J. Zhu. 2008. "Spatial Regression Models for Demographic Analysis." *Population Research and Policy Review* 27:17–42.
Chow, G.C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28:591–605.
Clark, D. and W. Hunter. 1992. "The Impact of Economic Opportunity, Amenities and Fiscal Factors on Age-Specific Migration Rates." *Journal of Regional Science* 32:349–65.
Clark, D. and C. Murphy. 1996. "Countywide Employment and Population Growth: An Analysis of the 1980s." *Journal of Regional Science* 36:235–56.
Commoner, B. 1972. "The Environmental Cost of Economic Growth." Pp. 339–63 in *Population, Resources and the Environment*. Washington, DC: Government Printing Office.
Cowen, D.J. and J.R. Jensen. 1998. "Extraction and Modeling of Urban Attributes Using Remote Sensing Technology." Pp. 164–88 in *People and Pixels: Linking Remote Sensing and Social Science,* edited by D. Liverman, E.F. Moran, R.R. Rindfuss, and P.C. Stern. Washington, DC: National Academy Press.
Cressie, N. 1993. *Statistics for Spatial Data*. New York: Wiley.
Deal, B., V.G. Pallathucheril, Z. Sun, J. Terstriep, and W. Hartel. 2005. "LEAM Technical Document: Overview of the LEAM Approach." Department of Urban and Regional Planning, University of Illinois at Urbana-Champaign.
Deller, S.C., T.H.S. Tsai, D.W. Marcouiller, and D.B.K. English. 2001. "The Role of Amenities and Quality-of-Life in Rural Economic Growth." *American Journal of Agricultural Economics* 83:352–65.
Dryden, I.L. and K.V. Mardia. 1998. *Statistical Shape Analysis*. Chichester, UK: John Wiley.
English, D.B.K., D.W. Marcouiller, and H.K. Cordell. 2000. "Tourism Dependence in Rural America: Estimates and Effects." *Society and Natural Resources* 13:185–202.
Espenshade, T.J. and J. Tayman. 1982. "Confidence Intervals for Postcensal State Population Estimates." *Demography* 19:191–210.
ESRI. 2000. ArcGIS ModelBuilder™ [Computer software]. Redlands, CA: Environmental Systems Research Institute.
Flowerdew, R. and M. Green. 1992. "Developments in Areal Interpolation Methods and GIS." *The Annals of Regional Science* 26:67–78.
Forrester, J.W. 1971. *World Dynamics*. Cambridge, MA: Wright-Allen Press.
Fotheringham, A.S. and D.W.S. Wong. 1991. "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis." *Environment and Planning A* 23:1025–34.
Frey, W.H. 1987. "Migration and Depopulation of the Metropolis: Regional Restructuring or Rural Renaissance?" *American Sociological Review* 52:240–57.
Friedman, S. and D.T. Lichter. 1998. "Spatial Inequality and Poverty Among American Children." *Population Research and Policy Review* 17:91–109.
Frisbie, W.P. and D.L. Poston. 1975. "Components of Sustenance Organization and Nonmetropolitan Population Change: A Human Ecological Investigation." *American Sociological Review* 40:773–84.
Fuguitt, G.V. 1985. "The Nonmetropolitan Population Turnaround." *Annual Review of Sociology* 11:259–80.
Fuguitt, G.V. and D. Brown. 1990. "Residential Preferences and Population Redistribution." *Demography* 27:589–600.
Garin, R.A. 1966. "A Matrix Formulation of the Lowry Model." *Journal of the American Institute of Planners* 32:361–66.

Graves, P. 1983. “Migration With a Composite Amenity.” *Journal of Regional Science* 23:541–46.
Graves, P. and P. Linneman. 1979. “Household Migration: Theoretical and Empirical Results.” *Journal of Urban Economics* 6:383–404.
Greenberg, M.R., D.A. Krueckeberg, and C.O. Michaelson. 1978. *Local Population and Employment Projection Techniques*. New Brunswick, NJ: Center for Urban Policy Research.
Hulten, C.R. and R.M. Schwab. 1984. “Regional Productivity Growth in U.S. Manufacturing, 1951–78.” *American Economic Review* 74:152–62.
Humphrey, C.R. 1980. “The Promotion of Growth in Small Urban Places and Its Impact on Population Change.” *Social Science Quarterly* 61:581–94.
Hunt, D. and D. Simmons. 1993. “Theory and Application of an Integrated Land-Use and Transport Modeling Framework.” *Environment and Planning B* 20:221–44.
Johnson, K.M. 1999. “The Rural Rebound.” *PRB Reports on America* 1(3):1–21.
Johnson, K.M. and R.L. Purdy. 1980. “Recent Nonmetropolitan Population Change in Fifty-Year Perspective.” *Demography* 17:57–70.
Keyfitz, N. 1972. “On Future Population.” *Journal of the American Statistical Association* 67:347–63.
———. 1982. “Can Knowledge Improve Forecasts?” *Population and Development Review* 8:729–51.
Knapp, T. and P. Graves. 1989. “On the Role of Amenities in Models of Migration and Regional Development.” *Journal of Regional Science* 29:71–87.
Koenker, R.W. and G.W. Bassett. 1982. “Robust Tests for Heteroscedasticity Based on Regression Quantiles.” *Econometrica* 50:43–61.
Land, K.C. 1986. “Methods for National Population Forecasts: A Review.” *Journal of the American Statistical Association* 81:888–901.
Land, K.C. and D. Cantor. 1983. “ARIMA Models of Seasonal Variation in U.S. Birth and Death Rates.” *Demography* 20:541–68.
Land, K.C. and S.H. Schneider. 1987. “Forecasting in the Social and Natural Sciences: Some Isomorphisms.” Pp. 7–31 in *Forecasting in the Social and Natural Sciences,* edited by K.C. Land and S.H. Schneider. Boston: D. Reidel.
Land Information & Computer Graphics Facility. 2000. “Mapping Growth Management Factors: A Practical Guide for Land Use Planning.” Land Information Bulletin Technical Paper No. 2. College of Agricultural and Life Sciences, University of Wisconsin–Madison.
Langford, M., D.J. Maguire, and D.J. Unwin. 1991. “The Areal Interpolation Problem: Estimating Population Using Remote Sensing in a GIS Framework.” Pp. 55–77 in *Handling Geographical Information: Methodology and Potential Applications,* edited by I. Masser and M. Blakemore. London: Longman Scientific & Technical.
Lewis, P.H. 1996. *Tomorrow by Design: A Regional Design Process for Sustainability*. New York: John Wiley & Sons.
Lowry, I.S. 1964. “A Model of Metropolis.” Memorandum RM-4035-RC. The Rand Corporation, Santa Monica, CA.
Lutz, W. 1994. *Population-Development-Environment: Understanding Their Interactions in Mauritius*. New York: Springer-Verlag.
Lutz, W. and J.R. Goldstein, Guest Eds. 2004. *Population Forecasting,* two special issues of *International Statistical Review* 72(1 and 2).
Lutz, W., J. Vaupel, and D. Ahlburg, Guest Eds. 1998. *Frontiers of Population Forecasting*, supplement to *Population and Development Review* 24.
Mandell, M. and J. Tayman. 1982. “Measuring Temporal Stability in Regression Models of Population Estimation.” *Demography* 19:135–46.
Marcouiller, D.W., K.K. Kim, and S.C. Deller. 2004. “Natural Amenities, Tourism and Income Distribution.” *Annals of Tourism Research* 31:1031–50.
Meadows, D.H., D.L. Meadows, and J. Randers. 1992. *Beyond the Limits: Confronting Global Collapse, Envisioning a Sustainable Future*. London: Earthscan Publications.
Meadows, D.H., D.L. Meadows, J. Randers, and W.I. Behrens. 1972. *The Limits to Growth*. New York: Potomac Associates.

Mincer, J. 1978. "Family Migration Decisions." *Journal of Political Economy* 86:749–73.

Morrison, C.J. and A.E. Schwartz. 1996. "State Infrastructure and Productive Performance." *American Economic Review* 86:1095–111.

Murdock, S., F. Leistritz, R. Hamm, S. Hwang, and B. Parpia. 1984. "An Assessment of the Accuracy of a Regional Economic-Demographic Projection Model." *Demography* 21:383–404.

Namboodiri, K. 1972. "On the Ratio-Correlation and Related Methods of Subnational Population Estimation." *Demography* 9:443–54.

O'Connor, J. 1989. "Political Economy of Ecology of Socialism and Capitalism." *Capitalism, Nature, Socialism* 3:93–106.

O'Neill, B.C. 2004. "Conditional Probabilistic Population Projections: An Application to Climate Change." *International Statistical Review* 72:167–84.

Pflaumer, P. 1992. "Forecasting U.S. Population Totals With the Box-Jenkins Approach." *International Journal of Forecasting* 8:329–38.

Pittenger, D.B. 1976. *Projecting State and Local Populations*. Cambridge, MA: Ballinger Publishing Company.

Porell, F. 1982. "Intermetropolitan Migration and Quality of Life." *Journal of Regional Science* 22:137–58.

Poston, D.L. and W.P. Frisbie. 1998. "Human Ecology, Sociology, and Demography." Pp. 27–50 in *Continuities in Sociological Human Ecology,* edited by M. Micklin and D.L. Poston. New York: Plenum Press.

Putman, S.H. 1991. *Integrated Urban Models II: New Research and Applications of Optimization and Dynamics*. London: Pion Limited.

Reibel, M. and M.E. Bufalino. 2005. "Street-Weighted Interpolation Techniques for Demographic Count Estimation in Incompatible Zone Systems." *Environment and Planning A* 37:127–39.

Riahi, K. and N. Nakicenovic, Guest Eds. 2007. *Greenhouse Gases—Integrated Assessment,* special issue of *Technological Forecasting and Social Change* 74(7).

Saboia, J. 1974. "Modeling and Forecasting Populations by Time Series: The Swedish Case." *Demography* 11:483–92.

Sanderson, W.C. 1995. "Predictability, Complexity, and Catastrophe in a Collapsible Model of Population, Development, and Environmental Interactions." *Mathematical Population Studies* 5:259–79.

———. 1998. "Knowledge Can Improve Forecasts: A Review of Selected Socioeconomic Population Projection Models." *Population and Development Review* 24(Suppl.):88–117.

Sanderson, W.C., S. Scherbov, B.C. O'Neill, and W. Lutz. 2004. "Conditional Probabilistic Population Forecasting." *International Statistical Review* 72:157–66.

Schmitt, R.C. 1953. "A New Method of Forecasting City Population." *Journal of the American Institute of Planners* 19:40–42.

———. 1954. "A Method of Projecting the Population of Census Tracts." *Journal of the American Institute of Planners* 20:102.

Schnaiberg, A. and K.A. Gould. 1994. *Environment and Society: The Enduring Conflict*. New York: St. Martin's Press.

Smith, S.K. 1987. "Tests of Forecast Accuracy and Bias for County Population Projections." *Journal of the American Statistical Association* 82:991–1003.

Smith, S.K., J. Nogel, and S. Cody. 2002. "A Regression Approach to Estimating the Average Number of Persons per Household." *Demography* 39:697–712.

Smith, S.K. and T. Sincich. 1992. "Evaluating the Forecast Accuracy and Bias of Alternative Population Projections for States." *International Journal of Forecasting* 8:495–508.

Smith, S.K. and J. Tayman. 2003. "An Evaluation of Population Projections by Age." *Demography* 40:741–57.

Smith, S.K., J. Tayman, and D.A. Swanson. 2001. *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

Southworth, F. 1995. "A Technical Review of Urban Land Use-Transportation as Tool for Evaluating Vehicle Travel Reduction Strategies." Report ORNL-6881. Oak Ridge National Laboratory, Oak Ridge, TN.

Stanbery, V.B. 1952. *Better Population Forecasting for Areas and Communities: A Guide Book for Those Who Make or Use Population Projections*. Washington, DC: U.S. Government Printing Office.

Swanson, D.A. and D.M. Beck. 1994. "A New Short-Term County Population Projection Method." *Journal of Economic and Social Measurement* 20:25–50.

Tayman, J. 1996. "The Accuracy of Small-Area Population Forecasts Based on a Spatial Interaction Land-Use Modeling System." *Journal of the American Planning Association* 62:85–98.

Tayman, J. and E. Schafer. 1985. "The Impact of Coefficient Drift and Measurement Error on the Accuracy of Ratio-Correlation Population Estimates." *Review of Regional Studies* 15(2):3–10.

Tayman, J., S.K. Smith, and J. Lin. 2007. "Precision, Bias, and Uncertainty for State Population Forecasts: An Exploratory Analysis of Time Series Models." *Population Research and Policy Review* 26:347–69.

Tayman, J. and D.A. Swanson. 1996. "On the Utility of Population Forecasts." *Demography* 33:523–28.

Tobler, W.R. 1979. "Smooth Pycnophylactic Interpolation for Geographic Regions." *Journal of the American Statistical Association* 74:519–30.

Trewartha, G. 1953. "A Case for Population Geography." *Annals of the Association of American Geographers* 43:71–97.

Wheeler, D. 1984. *Human Resource Policies, Economic Growth, and Demographic Change in Developing Countries*. Oxford: Clarendon Press.

Zuiches, J.J. and J.H. Rieger. 1978. "Size of Place Preferences and Life Cycle Migration: A Cohort Comparison." *Rural Sociology* 43:618–33.